jowua paper

11256-30952-1-SM

icst

Document Details

Submission ID

trn:oid:::3618:126092463

Submission Date

13 Jan 2026, 11:44 GMT+7

Download Date

13 Jan 2026, 11:46 GMT+7

File Name

11256-30952-1-SM.docx

File Size

2.6 MB

8 Pages

4,409 Words

26,904 Characters

21% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

62 Not Cited or Quoted 17%
Matches with neither in-text citation nor quotation marks

18 Missing Quotations 4%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

12% Internet sources
14% Publications
14% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **62** Not Cited or Quoted 17%
  Matches with neither in-text citation nor quotation marks
- **18** Missing Quotations 4%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

12% Internet sources
14% Publications
14% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	**Universitas Esa Unggul on 2026-01-12**	3%
2	Internet	**www.mdpi.com**	2%
3	Student papers	**Universitas Khairun on 2024-04-18**	1%
4	Internet	**digitallibrary.aau.ac.ae**	1%
5	Publication	**R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P...**	<1%
6	Internet	**download.bibis.ir**	<1%
7	Publication	**Zhanfeng Fan, Bo Chen, Zhijia Wang, Xiaojun Gou, Yutao Miao, Huajin Li, Zhishen...**	<1%
8	Publication	**"Proceedings of the TEPEN International Workshop on Fault Diagnostic and Prog...**	<1%
9	Publication	**Gaoli Zhao, Kefei Zhang, Liangzhi Wang, Wenyi Zhao, Weidong Zhang. "CIDNet: Cr...**	<1%
10	Publication	**Weiguo Ni. "Implementation of a Convolutional Neural Network (CNN)-based Obj...**	<1%

11 Internet
assets-eu.researchsquare.com <1%

12 Publication
Dan Huang, Guangyin Zhang, Zixu Li, Keying Liu, Wenguang Luo. "Light-YOLO: A l... <1%

13 Student papers
Sheffield Hallam University on 2026-01-08 <1%

14 Internet
hrcak.srce.hr <1%

15 Internet
ijece.iaescore.com <1%

16 Student papers
University of Essex on 2025-09-03 <1%

17 Publication
In Joo, Sunghoon Kim, Ginam Kim, Kwan-Hee Yoo. "SOIF-DN: Preserving Small Obj... <1%

18 Publication
N. Deluxni, Pradeep Sudhakaran. "Underwater debris detection using YOLOv12 w... <1%

19 Student papers
Xiamen University on 2024-12-07 <1%

20 Publication
Zhongle Ren, Zhe Du, Yu Zhang, Feng Sha, Weibin Li, Biao Hou. "Multi-Step Unsup... <1%

21 Publication
Falen Willem Ngongoloy, Yaulie D.Y. Rindengan, Arie S.M. Lumenta. "aplikasi pen... <1%

22 Publication
Shichang Liu, Xu Xu, Gwanggil Jeon, Junxin Chen, Ben-Guo He. "Deep learning bas... <1%

23 Publication
Arvind Dagur, Sohit Agarwal, Dhirendra Kumar Shukla, Shabir Ali, Sandhya Sharm... <1%

24 Student papers
Electrical & Computer Engineering on 2025-11-14 <1%

25 Student papers

Higher Education Commission Pakistan on 2023-09-15 <1%

26 Student papers

Imam Abdulrahman Bin Faisal University on 2025-12-19 <1%

27 Student papers

Indian Institute of Science Education and Research (IISER) Bhopal on 2023-03-03 <1%

28 Publication

Yang Tong, Hui Ye, Jishen Yang, Xiulong Yang. "ACD-DETR: Adaptive Cross-Scale D... <1%

29 Publication

Zhe Liu, Guiqing He, Yang Hu. "Multi-Level Contextual and Semantic Information ... <1%

30 Student papers

Ajou University Graduate School on 2023-06-20 <1%

31 Student papers

Al-Ibdaa Institute - Iraq on 2025-08-21 <1%

32 Student papers

City University on 2024-12-02 <1%

33 Publication

Fan Zhao, Xinlei Shao, Jiaqi Wang, Yijia Chen, Dianhan Xi, Yongying Liu, Jundong C... <1%

34 Publication

Goodness Oluchi Anyanwu, Cosmas Ifeanyi Nwakanma, Adinda Riztia Putri, Jae-M... <1%

35 Publication

Lei Du. "Object Detectors in Autonomous Vehicles: Analysis of Deep Learning Tec... <1%

36 Publication

Qiang He, Meng Li, Lianzhi Huo, Linlin Chen. "Learning to detect extreme objects ... <1%

37 Publication

Reagan Lewis, Teun Kostermans, Jan Wilhelm Brovold, Talha Laique, Marko Ocep... <1%

38 Publication

Rong Ye, Guoqi Shao, Ziyi Yang, Yuchen Sun, Quan Gao, Tong Li. "Detection Model... <1%

39 Student papers

Singapore Institute of Technology on 2023-07-27 <1%

40 Publication

Suneeta Satpathy, Álvaro Rocha, Sachi Nandan Mohanty, Tanupriya Choudhury. "... <1%

41 Student papers

University Tun Hussein Onn Malaysia on 2025-05-06 <1%

42 Publication

Zhonghua Li, Biao Hou, Zitong Wu, Bo Ren, Chen Yang. "FCOSR: A Simple Anchor-... <1%

43 Internet

arxiv.org <1%

44 Internet

koreascience.or.kr <1%

45 Internet

mdpi-res.com <1%

46 Publication

Mengran Yang, Shuangshuang Han. "ETS-YOLO: An Efficient YOLO-based Model f... <1%

47 Student papers

University of Glasgow on 2024-04-24 <1%

48 Student papers

Association of Educators on 2023-10-30 <1%

49 Publication

Shuangbao Ma, Wennan Li, Li Wan, Guoqin Zhang. "A Lightweight Fire Detection ... <1%

50 Student papers

University of Lancaster on 2023-09-07 <1%

51 Student papers

University of Southampton on 2025-05-06 <1%

JIKO (Jurnal Informatika dan Komputer) Accredited KEMDIKTISAINTEK, No.0173/C3/DT.05.00/2025
Vol. x, No. x, April 2025, pp. x-x p-ISSN: 2614-8897
DOI: 10.33387/jiko e-ISSN: 2656-1948

LIGHTWEIGHT DOLPHIN DETECTION USING EFFICIENT NETWORK IN AQUATIC ENVIRONMENTS

Febriyanti Ludja[1], Florensce Sumarauw[2], Robby Moody Lintong[3], Steven R. Sentinuwo[4], Alwin M. Sambul[5], Muhamad Dwisnanto Putro[6]
[1,2,3,4] Master Program of Informatics, Faculty of Engineering, Sam Ratulangi University, Manado, 95115, Indonesia
Email: *[1]febriyantiludja026@student.unsrat.ac.id, [2]florenscesumarauw026@student.unsrat.ac.id, [3]robbymoodylintong026@student.unsrat.ac.id, [4]steven@unsrat.ac.id, [5]asambul@unsrat.ac.id, [6]dwisnantoputro@unsrat.ac.id

(Received: dd mmm yyyy, Revised: dd mmm yyyy, Accepted: dd mmm yyyy)

Abstract

Dolphin monitoring plays a crucial role in maintaining the balance of the marine ecosystem and is an attraction for ecotourism, accurate detection is often hampered by light refraction, water turbidity, and the rapid movement of dolphins. This study proposes the application of YOLO11-Nano as a lightweight detection architecture for real-time monitoring on resource-constrained devices, which has been rarely explored in aquatic scenarios. The model is optimised to balance inference speed and detection accuracy, remaining competitive in various environmental conditions. Experimental results show that YOLO11-Nano achieves 6.4 GFLOPs, 2.59 million parameters, 65.0% mAP50, 43.1% mAP50:95, and 18.34 FPS, confirming stable performance under various lighting and underwater noise conditions. This indicates the potential of YOLO11-Nano for application in real-time water monitoring systems on devices with limited resources, supporting automated dolphin detection as part of marine ecosystem conservation efforts.

Keywords: *Deep Learning, Dolphin Detection, Real-time, YOLO11-Nano*

**Corresponding Author:* Muhamad Dwisnanto Putro*

1. INTRODUCTION

Underwater object detection is a challenging problem in computer vision due to factors such as water turbidity, light variation, surface reflections, and particles that reduce image quality [1], [2]. These conditions make object identification much more difficult than detection in terrestrial environments. Therefore, detection methods that can adapt to dynamic water conditions are essential for achieving accurate and efficient results. Recently, underwater optical image-based object detection methods have become popular as an efficient approach [3]. However, this task remains very challenging due to the complex underwater environment and lighting conditions [4].

Among underwater species, dolphins play an important ecological role and serve as indicators of marine ecosystem health. Accurate detection of these mammals supports population monitoring, migration tracking, and conservation management, contributing to a deeper understanding of marine biodiversity [5]. However, findings from long-term monitoring studies indicate a sustained decline in dolphin populations in recent years [6], [7]. On the other hand, the behavioural characteristics of these mammals, such as their rapid movement, frequent surfacing and diving, and varied body orientation in different lighting conditions, as well as the degradation of underwater image quality and complex backgrounds, make detection challenging [8]. Traditional observation methods, such as manual visual surveys and acoustic monitoring, are time-consuming and often unreliable. These limitations have driven the development of automated vision-based systems. The integration of deep learning into these systems enables real-time dolphin detection, even in complex aquatic environments. In particular, lightweight architectures such as YOLO11-Nano offer real-time detection with high accuracy and efficiency, making them suitable for integration on mobile platforms with limited resources. Studies [9] show that YOLOv11-Nano has a high degree of suitability for drone-based applications, particularly in supporting real-time processing and inference efficiency. These characteristics make YOLO11-Nano a promising approach for automated marine monitoring and support the acceleration of marine ecosystem conservation efforts[10].

Traditional observation methods, such as manual visual surveys and acoustic monitoring, are time-consuming and often unreliable. These limitations have driven the development of automated vision-based systems. The integration of deep learning into these systems enables real-time detection of dolphins, even in complex aquatic environments. In particular, lightweight architectures such as YOLO11-Nano offer real-time detection with high accuracy and efficiency, even in complex underwater conditions, thereby supporting automated marine monitoring and accelerating conservation efforts [11].

Deep learning is rapidly advancing the field of underwater object detection. As a field of computer vision, object detection focuses on identifying and locating objects in images or videos [8]. In addition to classification, this method provides spatial location using bounding boxes. However, underwater detection remains challenging due to factors such as lighting variations, water turbidity, and complex backgrounds [12]. Deep learning-based methods help overcome these challenges by automatically extracting discriminative features from image data, improving detection accuracy in aquatic environments [13].

Convolutional Neural Networks (CNNs) are now a cornerstone of modern object detection, primarily due to *their* ability to extract representative features from images [3]. Several studies *have demonstrated* that CNNs are *highly effective for* anomaly detection [14]. This concept applies directly to dolphin detection: against the complex sea background, the presence of a dolphin *is, in essence, an anomaly*. Therefore, the CNN approach can improve the accuracy and reliability of dolphin detection systems in aquatic environments. One algorithm from CNN development is You Only Look Once (YOLO), which integrates classification and localization in one step, enabling real-time detection and efficiency [15]. This system continues to evolve by offering improvements in accuracy, processing speed, and the ability to detect small objects, making it highly relevant for detecting objects such as dolphins.

This architecture presents lightweight variants such as YOLO11-Nano, which are optimized to maintain a balance between accuracy and speed while requiring low computational complexity, making them suitable for deployment on resource-constrained devices [16], [9]. Nevertheless, underwater object detection systems continue to face significant challenges in the field of computer vision due to low visibility, poor lighting conditions, and water turbidity, which often lead to detection errors [17], [18]. To address these issues, various deep learning models based on CNNs and YOLO architectures have been widely adopted because of their ability to balance detection speed and accuracy [19]. However, conventional YOLO architectures do not always exhibit stable performance under diverse and dynamically complex underwater environmental conditions. To overcome these limitations, several studies have proposed architectural modifications to enhance feature extraction capabilities, including the integration of Convolutional Block Attention Modules (CBAM) and Transformer modules into YOLOv5s, as reported in [20]. Although these approaches have been shown to improve detection accuracy, the introduction of additional components generally increases computational complexity while reducing inference efficiency, thereby limiting their applicability in real-time systems operating on resource-constrained devices. In the context of dolphin detection, previous studies [21] have focused on enhancing detection performance through architectural adaptations of lightweight YOLO architectures to accommodate underwater environmental characteristics and the rapid dynamics of object motion. While these approaches demonstrated performance improvements, they remain reliant on earlier generations of YOLO architectures and do not provide a comprehensive evaluation of the performance of the latest YOLO generations. In particular, evaluations of YOLO11-Nano as a lightweight detector for fast-moving marine mammals, such as dolphins, in dynamic underwater environments remain limited.

This research contributes by presenting a systematic experimental evaluation of the performance of the YOLO11-Nano architecture for dolphin detection in underwater environments. The study specifically examines the ability of YOLO11-Nano to cope with dynamic water conditions, including variations in lighting, turbidity levels, and background complexity, which represent real challenges in marine monitoring scenarios. This analysis provides empirical insights into the potential and limitations of YOLO11-Nano as a lightweight detector for real-time monitoring of marine mammals in resource-constrained systems.

2. RESEARCH METHOD

This study proposes using YOLO11-Nano as a solution for real-time dolphin detection in complex aquatic environments, due to its high computational efficiency while maintaining competitive detection accuracy, making it ideal for implementation on devices with limited resources. This architecture is a lightweight YOLO family variant designed to balance inference speed and detection accuracy. As shown in Figure 1, it consists of three main components: Backbone, Neck, and Head.

2.1 Backbone

The backbone section in the YOLO11-Nano architecture serves as the primary component in the feature extraction process, responsible for learning the visual representation of input images. This backbone comprises multiple fundamental blocks arranged sequentially to achieve a balance between

JIKO (Jurnal Informatika dan Komputer) Accredited KEMDIKTISAINTEK, No.0173/C3/DT.05.00/2025
Vol. x, No. x, April 2025, pp. x-x p-ISSN: 2614-8897
DOI: 10.33387/jiko e-ISSN: 2656-1948

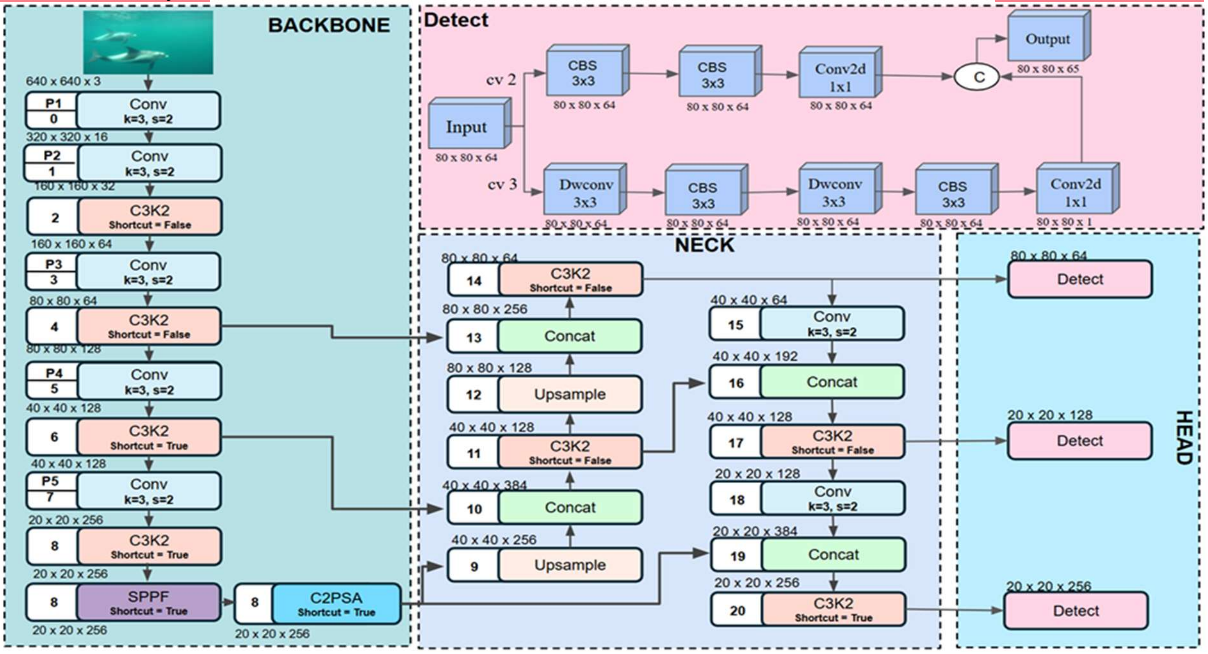

Figure 1. Architecture of YOLO11-Nano

computational efficiency and feature representation capabilities. Each block is designed to progressively enrich visual information, resulting in increasingly discriminative feature maps. The main blocks that form the backbone include:

2.1.1 C3K2

The C3K2 module, as illustrated in Figure 2, begins with a 1 × 1 convolution that serves to adjust the channel dimensions, where some channels are used for feature extraction, and the rest are retained as identity connections to improve computational efficiency. This module supports two operational configurations: the C3K configuration, which consists of three convolutional layers with flexible kernel sizes, and the C2f configuration, which uses two 3 × 3 convolutional layers. The output from each C3K branch or bottleneck is then combined with identity features and further processed through 1 × 1 convolution to strengthen inter-channel interactions. This concise and flexible design enables C3K2 to produce rich and robust feature representations, making it suitable for computer vision tasks in underwater environments.

2.1.2 SPPF

The Spatial Pyramid Pooling-Fast (SPPF) module is applied after the last C3K2 module in the backbone architecture to enrich feature representation, as illustrated in Figure 3. This process begins with a 1 × 1 convolution that serves to compress the channel dimensions so that the number of channels can be adjusted. Next, the resulting features are processed through three consecutive 5 × 5 MaxPooling layers, where each stage preserves the spatial dimensions while gradually expanding the receptive field. This approach allows the model to capture local and global information effectively without significantly increasing the computational load. The output from each pooling stage is then combined and reprocessed using 1 × 1 convolution to integrate multi-scale features into the final feature map. This design enables the SPPF module to consistently extract feature representations at various scales, thereby contributing to improved detection accuracy without adding significant computational complexity

2.1.3 C2PSA

In the final stage of the backbone, YOLO11-Nano integrates the C2PSA module after the SPPF layer to enhance feature representation quality, as shown in Figure 4. This module begins with a 1 × 1 convolution to adjust the channel dimensions, followed by feature partitioning into multiple parallel branches. One branch is processed through a Partial Self-Attention (PSA) block [22] to model global relationships between features through a self-attention mechanism, and the other branch retains the original features to maintain representation efficiency. The outputs from all branches are then concatenated and processed through a 1 × 1 convolution at the final stage to integrate inter-channel information, resulting in a more discriminative feature map before being passed on to the next stage in the network.

2.2 Neck

The Neck section in the YOLO11-Nano architecture integrates feature information from various scales extracted by the backbone. This process involves up-sampling and down-sampling operations to align the dimensions of the feature map,

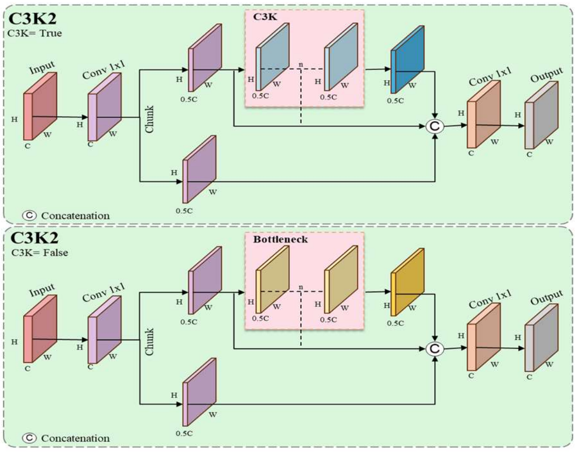

Figure 2. C3K2 module. The true configuration employs C3K blocks, whereas the False configuration utilizes bottleneck blocks.

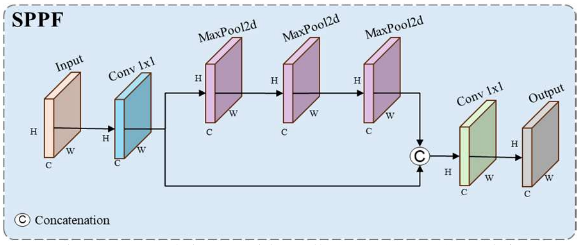

Figure 3. Spatial Pyramid Pooling Fast (SPPF) module

thereby strengthening the connections between elements at different convolution stages. This module uses a combination of Path Aggregation Network (PAN) [23] and Feature Pyramid Network (FPN) [24] structures to generate three feature levels using bottom-up and top-down path aggregation strategies. The model integrates C3K2 blocks to capture deeper contextual information while combining features from multiple receptive fields, thereby improving multiscale feature representation without increasing computational complexity, and CBS, a combination of convolution, batch normalization, and SiLU activation function to maintain stable data distribution and improve the model's ability to recognize complex patterns. The aggregated information is combined with features extracted from the backbone, strengthening the combined representation before it is forwarded to the detection head.

2.3 Head

The detection system in YOLO11-Nano generates prediction outputs in the Head section, which serves as the final stage for producing object detection results based on features processed by the backbone and neck. In a single inference step, this component performs two main tasks simultaneously: object classification and bounding box regression. The Head network operates on different feature map scales of 80×80, 40×40, and 20×20 to robustly detect objects of varying sizes. This architecture uses three detection layers, each designed for a specific scale: large feature maps detect small objects, small feature

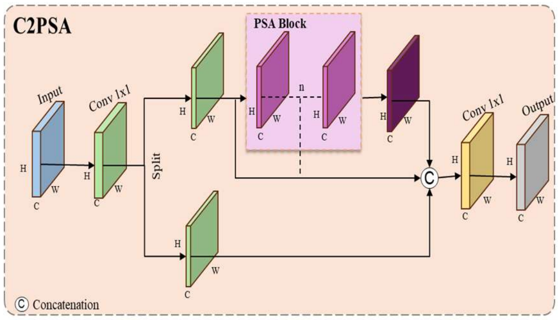

Figure 4. Cross-Stage Partial Parallel Split Attention (C2PSA) module.

maps handle large objects, and medium feature maps target medium-sized objects. Each detection layer consists of two parallel branches formed by successive 3×3 convolution layers followed by 1×1 convolution layers. The regression branch predicts the bounding box coordinates (x, y, w, h), while the classification branch estimates the probability of the detected object class.

This multi-scale detection strategy enables YOLO11-Nano to achieve high adaptability and robustness when detecting objects at different spatial resolutions. The optimization process utilizes a combination of three loss functions: Complete Intersection over Union (CIoU) Loss to maximize localization accuracy by considering the overlapping area, center distance, and aspect ratio [25]; Distribution Focal Loss (DFL) to improve bounding box regression by modeling localization targets as probability distributions [26], and Binary Cross-Entropy (BCE) Loss to optimize object prediction and class probability estimation by minimizing the divergence between predicted labels and actual labels. The integration of these loss functions ensures a balance between localization precision and classification accuracy, enabling YOLO11-Nano to maintain optimal detection performance on objects of varying scales and complexities.

2.4 Implementation Setup

To evaluate the effectiveness of the proposed method, experiments were conducted to achieve optimal performance while maintaining a balance between computational efficiency and model accuracy. The system configuration and dataset were carefully designed to align with the research objectives and support the model training process optimally. Further implementation details will be explained below.

2.4.1 Training and Testing Configuration

In this experiment, training and testing of the YOLO11-Nano model were performed using the configuration summarized in Table 1. Training was conducted on the Kaggle platform with an NVIDIA Tesla P100 GPU, which provides high computing

Table 1. Training and Testing Configuration

Parameters	Setup
Platfrom/device	Kaggle
GPU	P100
Image Size	640 x 640 pixels
Epochs	300
Batch Size	32
Optimizer	Stochastic Gradient Descent (SGD)
Learning Rate	0,01

power for efficient training and inference. The input image size was 640 × 640 pixels to balance spatial detail accuracy and computational efficiency. The model was trained for 300 epochs with a batch size of 32, ensuring stable convergence while effectively utilizing GPU memory. Optimization was performed using the Stochastic Gradient Descent (SGD) algorithm due to its robustness and consistent convergence in object detection tasks. A learning rate set at 0.01 was applied to enhance convergence and mitigate overfitting, thereby improving generalization performance.

For inference, model testing, and evaluation were performed on a local device using a CPU configuration. This approach was adopted to examine how well the model could be applied in a real-world implementation environment with limited computational resources.

2.4.1 Dataset

The study employed the dolphin dataset, which was obtained from previous research [21]. This dataset consists of 5,493 images, including 4,122 images for training, 822 samples for testing, and 549 images for validation, as shown in Table 2. Data augmentation was applied exclusively to the training set to improve model generalization, while the validation and testing sets remained unchanged to ensure objective performance evaluation.

Table 2. Dataset Configuration

Parameters	Setup
Training Data	4.122 images
Validation Data	822 images
Testing Data	549 images

This data splitting strategy resulted in a representative and balanced dataset, supporting robust model training and reliable performance assessment. Every sample in the dataset was labeled with bounding boxes to identify and localize dolphins.

The dataset contains both above-water and underwater images, featuring diverse lighting conditions, ocean backgrounds, and challenging detection scenarios such as water reflections and dynamic dolphin movements, as shown in Figure 5. Overall, the dataset contains 5,493 images with approximately 4,900 labeled dolphin instances. Each instance represents the presence of a dolphin in an image, either as a single subject or as part of a group.

Figure 5. The dolphin dataset captured under two conditions, above water and underwater

3. RESULTS AND DISCUSSION

The performance of the proposed YOLO11-Nano model was evaluated based on detection accuracy, computational efficiency, robustness to various aquatic environmental conditions, and execution time efficiency. The evaluation was conducted using a dolphin dataset, with variations in lighting conditions, turbidity levels, and light reflection on the water surface.

3. 1 Evaluation on Dataset

The detection accuracy of the proposed model was evaluated using standard evaluation metrics, namely mean Average Precision at an IoU threshold of 0.5 (mAP50) and mean Average Precision in the IoU range from 0.5 to 0.95 (mAP50:95). As shown in Table 3, the YOLO11-Nano model achieved an mAP50 value of 65.0% and an mAP50:95 value of 43.1%, with 2.59 million parameters and a computational complexity of 6.4 GFLOPs. This performance is compared to other lightweight detectors in the YOLO family, namely YOLOv12-Nano, which achieved an mAP50 of 60.8%, mAP50:95 of 41.9%, 2.56 million parameters, and a complexity of 6.5 GFLOPs, and YOLOv10-Nano, which achieved mAP50 of 59.6%, mAP50:95 of 39.9%, 2.70 million parameters, and a complexity of 8.4 GFLOPs. These results show that YOLO11-Nano is able to achieve an excellent balance between accuracy and efficiency, surpassing the performance of YOLOv10 and YOLOv12 in terms of detection precision and computational cost, while maintaining a lightweight architecture that is easy to implement on devices with limited resources.

Table 2. Performance and Efficiency Results

Model	GFLOPs	Parameter	mAP 50%	mAP 50-95%
YOLOv12-Nano	6.5	2.56	60.8	41.9
YOLO11-Nano	**6.4**	**2.59**	**65.0**	**43.1**
YOLOv10-Nano	8.4	2.70	59.6	39.9

JIKO (Jurnal Informatika dan Komputer) Accredited KEMDIKTISAINTEK, No.0173/C3/DT.05.00/2025
Vol. x, No. x, April 2025, pp. x-x p-ISSN: 2614-8897
DOI: 10.33387/jiko e-ISSN: 2656-1948

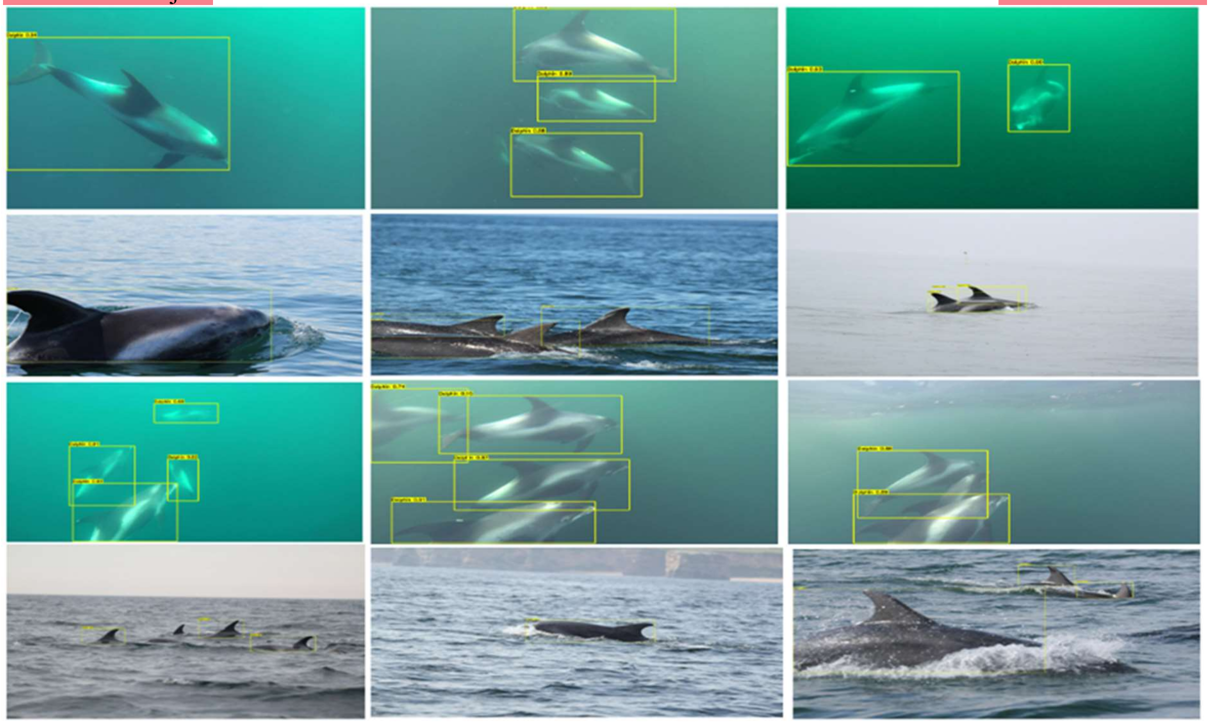

Figure 6. Dolphin detection results using the YOLO11-Nano model

Figure 6 shows the results of dolphin detection using the YOLO11-Nano model in above- and below-water conditions. In general, the model is able to detect dolphins accurately, even when the object is only partially visible, is at a depth with low visibility, or is affected by surface light reflections. This demonstrates the model's ability to generalize and deal with variations in complex aquatic environmental conditions. However, some detection errors were still found. In some images, dolphins were detected more than once (multiple bounding boxes), while in other cases, some individuals were only partially detected or not detected at all. These conditions indicate that although the model has achieved a high level of precision, performance improvements are still needed so that detection accuracy can be further improved and detection errors can be minimized in the future.

To further evaluate the classification performance of the proposed YOLO11-Nano model, a normalised confusion matrix is shown in Figure 7. The test results indicate that the model has a reliable ability to distinguish dolphin objects from background classes. In the dolphin category, the model produced a true prediction value of 0.56, indicating that dolphin features can be well recognised in the test dataset. In addition, the background class is identified with a value of 1.00, which suggests that the model is consistently able to recognise non-dolphin areas without classification errors. These results indicate that the YOLO11-Nano model can effectively extract the distinguishing characteristics between target objects and the background, thus supporting stable classification performance in underwater detection scenarios.

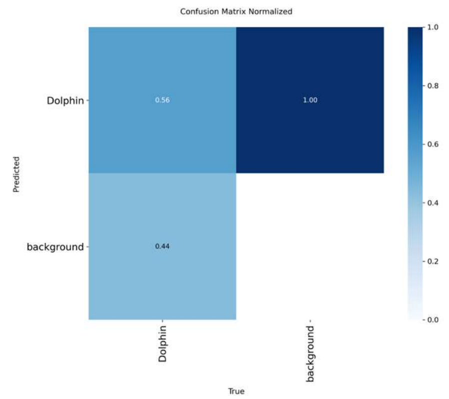

Figure 7. Normalized confusion matrix of the proposed YOLO11-Nano model on the test dataset.

3. 2 Runtime Efficiency

The evaluation of execution time efficiency focused on measuring the inference performance and computational cost of the YOLO11-Nano model in real-time detection tasks. As shown in Table 4, this model achieved the highest inference speed of 18.34 FPS with a computational complexity of 6.4 GFLOPs, surpassing other lightweight models such as YOLOv12-Nano (16.96 FPS, 6.5 GFLOPs) and YOLOv10-Nano (17.63 FPS, 8.4 GFLOPs). These results show that YOLO11-Nano has superior execution time efficiency with low computational requirements, making it very suitable for implementation on devices with limited resources in real-time detection applications.

The increase in YOLO11-Nano inference speed is due to its optimized architectural design,

Table 4. Runtime Efficiency Comparison of YOLO-Nano Models

Model	GFLOPs	FPS
YOLOv12-Nano	6.5	16.96
YOLO11-Nano	**6.4**	**18.34**
YOLOv10-Nano	8.4	17.63

specifically the efficient combination of C3K2 and C2PSA modules that improve feature extraction while reducing redundant calculations. This balance between speed and accuracy ensures that the model not only excels in detection precision but also meets the practical needs for real-time dolphin monitoring in aquatic environments.

4. CONCLUSION

This study demonstrates the effectiveness of the YOLO11-Nano architecture to perform real-time dolphin detection in complex aquatic environments. This model achieves a mean Average Precision mAP50 of 65.0% and mAP50:95 of 43.1%, with 2.59 million parameters and a computational complexity of 6.4 GFLOPs, demonstrating its ability to achieve a balance between detection accuracy and computational efficiency. Compared to other variants such as YOLOv10-Nano and YOLOv12-Nano, the proposed model outperforms in terms of both accuracy and inference speed, attaining 31.40 FPS, confirming its potential for real-time application on devices with limited resources. As a major contribution, this study presents a comprehensive evaluation of the YOLO11-Nano architecture for underwater dolphin detection, highlighting its effectiveness as a lightweight yet accurate detection framework.

Furthermore, experimental results show that YOLO11-Nano is capable of overcoming various underwater challenges, including low visibility conditions, lighting variations, and surface light reflections. However, some detection inaccuracies were still found, such as overlapping bounding boxes or partial detection. Therefore, further research could focus on improving the attention mechanism to increase precision. Overall, this research makes an important contribution to the development of lightweight deep learning models for marine species monitoring, while providing a strong foundation for the development of automated and efficient aquatic ecosystem monitoring systems.

Acknowledgment

The authors would like to acknowledge the **AIVISION** team for their valuable contributions and support throughout this study. Their expertise and insights in the field of computer vision and deep learning, along with the availability of computational facilities and technical assistance, played an important role in supporting the experimental process and the preparation of this manuscript.

5. REFERENCE

[1] A. Homoud, S. Das, and S. Townley, 'Challenges in underwater object detection and video segmentation using deep learning', in *2024 First International Conference for Women in Computing (InCoWoCo)*, 2024, pp. 1–6..

[2] Y. Zhang and X. Zhang, 'YOLO-AES A Lightweight Model for Real-Time Underwater Object Detection', in *2024 5th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, 2024, pp. 611–615.

[3] M. Elmezain, L. S. Saoud, A. Sultan, M. Heshmat, L. Seneviratne, and I. Hussain, 'Advancing underwater vision: a survey of deep learning models for underwater object recognition and tracking', *IEEE Access*, 2025.

[4] L. Chen *et al.*, 'Underwater Optical Object Detection in the Era of Artificial Intelligence: Current, Challenge, and Future', *ACM Computing Surveys*, 2025.

[5] M. D. Putro, Y. Mose, A. C. Andaria, J. Litouw, V. C. Poekoel, and X. Najoan, 'Streamlining Deep Learning Network for Real-time Sea Turtle Detection', *Jurnal Rekayasa Elektrika Vol*, vol. 20, no. 3, pp. 116–124, 2024.

[6] S. B. Kurniawan et al., "Tackling marine pollution in the blue economy: Synergies between wastewater treatment technologies and governmental policies," Marine Pollution Bulletin, vol. 212, p. 117431, Jan. 2025, doi: 10.1016/j.marpolbul.2024.117431.

[7] A. March, M. Bennett, M. Germishuizen, T. Evans, and P. Failler, "The status of Blue Economy development in Africa," Marine Policy, vol. 165, p. 106205, Jul. 2024, doi: 10.1016/j.marpol.2024.106205.

[8] H. Zhang, Q. Zhang, P. A. Nguyen, V. C. S. Lee, and A. Chan, 'Chinese white dolphin detection in the wild', in *Proceedings of the 3rd ACM International Conference on Multimedia in Asia*, 2021, pp. 1–5.

[9] M. F. Riftiarrasyid, F. L. Gaol, H. Soeparno, and Y. Arifin, 'Suitability of Latest Version of YOLOv11 in Drone Development Studies', in 2024 7th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2024, pp. 504–509.

[10] S. Cheng, Y. Han, Z. Wang, S. Liu, B. Yang, and J. Li, 'An underwater object recognition system based on improved yolov11', *Electronics*, vol. 14, no. 1, p. 201, 2025.

[11] R. Maglietta, R. Carlucci, C. Fanizza, and G. Dimauro, 'Machine learning and image processing methods for cetacean photo identification: a systematic review', *IEEE Access*, vol. 10, pp. 80195–80207, 2022.

[12] W. Ouyang and Y. Wei, 'An anchor-free detector with channel-based prior and bottom-enhancement for underwater object detection', *IEEE Sensors Journal*, vol. 23, no. 20, pp. 24800–24811, 2023.

[13] M. D. Putro, A. Sutrisno, I. S. Manembu, I. Y. Chun, and T.-H. Oh, 'STAR: Sea Turtle Basic Activity Recognizer Network Via Efficient Transformer', *IEEE Access*, 2025.

[14] M. Jain and A. Shah, 'Anomaly Detection Using Convolutional Neural Networks (CNN)', *ESP International Journal of Advancements in Computational Technology* (ESP-IJACT), vol. 2, no. 3, pp. 12–22, 2024.

[15] T. Diwan, G. Anirudh, and J. V. Tembhurne, 'Object detection using YOLO: challenges, architectural successors, datasets and applications', *multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, 2023.

[16] R. Khanam and M. Hussain, 'Yolov11: An overview of the key architectural enhancements', *arXiv preprint arXiv*:2410. 17725, 2024.

[17] V. Mane, S. Patwardhan, P. Pethkar, and R. Patil, 'Underwater object tracking and classification of marine animals', in 2024 *International Conference on Inventive Computation Technologies* (ICICT), 2024, pp. 1054–1058.

[18] P. Vijayalakshmi, M. Seetharaman, and E. Praveen, 'Underwater Image Enhancement and Object Recognition Using CNN Algorithm', in 2024 5th *IEEE Global Conference for Advancement in Technology* (GCAT), 2024, pp. 1–5.

[19] [20] C. Prathima, C. Silpa, A. Charitha, G. Harshitha, C. S. Charan, and G. R. Sailendra, 'Detecting and Recognizing Marine Animals Using Advanced Deep Learning Models', in 2024 *International Conference on Expert Clouds and Applications* (ICOECA), 2024, pp. 950–955.

[20] X. Xu, J. Hu, J. Yang, Y. Ran, and Z. Tan, 'A fish detection and tracking method based on improved inter-frame difference and YOLO-CTS', *IEEE Transactions on Instrumentation and Measurement*, 2024.

[21] F. Ludja, R. M. Lintong, F. Sumarauw, A. M. Sambul, S. R. Sentinuwo, and M. D. Putro, 'DOLPHIN DETECTION USING AN ENHANCED LIGHTWEIGHT YOLO ARCHITECTURE', JIPI (*Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*), vol. 10, no. 3, pp. 2874–2884, 2025.

[22] H. Zhang, K. Zu, J. Lu, Y. Zou, and D. Meng, 'EPSANet: An efficient pyramid squeeze attention block on convolutional neural network', in *Proceedings of the asian conference on computer vision*, 2022, pp. 1161–1177.

[23] Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.

[24] Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.

[25] Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Yang, J. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. Adv. Neural Inf. Process. Syst. 2020, 33, 21002–21012..

[26] Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. Proc. AAAI Conf. *Artif. Intell*. 2020, 34, 12993–13000. [CrossRef]